Civil and Environmental Research www.iiste.org
ISSN 2224-5790 (Paper) ISSN 2225-0514 (Online)
Vol.6, No.6, 2014

Fitting Statistical Distributions of Monthly Rainfall for Some Iraqi Stations

Najm Obaid Salim Alghazali [1*] Dhelal Adnan Hussein Alawadi [2]

1. Asst. Prof. Doctor, Civil Engineering Department, BabylonUniversity, Iraq
2. M.Sc. Student, Civil Engineering Department, Babylon University, Iraq

* E-mail of the corresponding author: dr.nalghazali@gmail.com

Abstract

In this study three statistical distributions were fitted to thirteen Iraqi stations of monthly rainfall observations: Mosul, Kirkuk, Khanaqin, Ramadi, Baghdad, Karbala, Hilla, Najaf, Diwaniya, Samawa, Nasiriyah, Amara and Basrah for the period (1970-2010) for all stations except Ramadi (1981-2010) and Hilla (1980-2010). These distributions were: Normal, Gamma and Weibull distributions. Method of moments was used for estimating parameters and two goodness of fit test were used: Chi-Square and Kolmogorov-Smirnov tests. Chi-Square test showed that Gamma distribution was suitable for five stations which were Ramadi, Baghdad, Hilla, Najaf and Samawa stations while, Normal and Weibull distributions were not suitable for any station. Kolmogorov-Smirnov test showed that none of the three distributions was suitable for any station.

Keywords: Rainfall in Iraq, Statistical distributions, Normal distribution, Gamma distribution, Weibull distribution

1. Introduction

The amount of rainfall received over an area is an important factor in assessing the amount of water available to meet the various demands of agriculture, industry, and other human activities. Therefore, the study of the distribution of rainfall in time and space is very important for the welfare of the national economy. Many applications of rainfall data are enhanced by knowledge of the actual distribution of rainfall rather than relying on simple summary statistics.

Rainfall data from arid and semi-arid regions are best fit to one of several probability distribution functions such as Normal (Gaussian), Log-Normal, Gamma, Weibull, and Gumbel distributions (Maliva and Missimer, 2012).

Ozturk (1981) reviewed some of probability distribution models for precipitation totals and their applications. The general properties of a probability distribution model which was a mixture of Gamma and Poisson distribution were discussed. A new approximation was given for the solution of likelihood equations and the efficiency of the estimators was obtained. An application was also made using monthly precipitation totals.

Sen and Eljadid (1999) studied the rainfall distribution function for Libya and rainfall prediction. Libyan monthly rainfall distributions were found to abide by Gamma probability distribution function which was confirmed on the basis of chi-square tests. Almost all the rainfall sequences recorded for at least the last 20 years in Libya were investigated statistically and Gamma distribution parameters were calculated at existing stations. The shape and scale parameters were then regionalized. Predictions of 10, 25, 50 and 100 mm rainfall amounts were achieved by this probability function.

Al-Mansory (2005) identified the proper theoretical statistical distributions for the data set of highest monthly rainfall for Basrah station, south of Iraq, for 75-given years with missing data. Six distributions were selected: Normal, Log-Normal, Log-Normal type III, Pearson type III, Log-Person type III, and Gumbel. Maximum likelihood method was selected to estimate theoretical distribution parameters. Adequacy test was conducted using Chi-square test. Results indicated that Person type III and Gumbel distributions were the proper for describing maximum monthly rainfall in the area being examined.

Husak et al. (2006) studied the use of Gamma distribution to represent monthly rainfall in Africa for drought monitoring applications. They demonstrated the feasibility of fitting cell-by-cell probability distributions to grids of monthly interpolated, continent-wide data Gamma distribution was well suited to these applications because it was fairly familiar to African scientists, and capable of representing a variety of distribution shapes. This study tested the goodness-of-fit using the Kolmogorov–Smirnov test, and compared these results against Weibull distribution which was commonly used in rainfall events. They found that Gamma distribution was suitable for roughly 98% of the locations over all months. The techniques and results presented in this study provided a foundation for use of the Gamma distribution to generate drivers for various rain related models.

Olumide et al. (2013) fitted various probability distribution models to various rainfall and runoff for the Tagwai dam site in Minna, Niger State, Nigeria to evaluate the model that was best suitable for the prediction of their values and subsequently using the best model to predict for both the expected yearly maximum daily-rainfall and yearly maximum daily-runoff at some specific return periods. The distributions Gumbel, log Gumbel, Normal and Log-Normal distribution were used in the study. The Normal distribution model was found most appropriate for the prediction of yearly maximum daily-rainfall and the Log-Gumbel distribution model was the most appropriate for the prediction of yearly maximum daily-runoff.

2. Probability Distributions of Rainfall Data

Analysis of rainfall data strongly depends on its distribution pattern. It has long been a topic of interest in the fields of meteorology in establishing a probability distribution that provides a good fit to rainfall. Papalexiou et al. (2012) proposed a typical procedure for selecting a distribution law of rainfall with three steps:

1. Select a priori some of the many parametric families of distributions.
2. Estimate the parameters according to one of the many fitting methods.
3. Choose the one best fitted according to some metric or fitting test.

In this study three probability distributions are selected, i.e., Normal, Gamma (2 Parameters) and Weibull (2 Parameters) distributions.

2.1 Normal Distribution

The most important probability distribution in theory as well as in application is the *Gaussian* or *normal* distribution. Its corresponding probability distribution function (PDF) is:

$$F(x) = P(X \leq x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-(x-\mu)^2/2\sigma^2} \, .dx \quad (1)$$

where :
μ : the mean
σ : the standard deviation
$x_1, x_2, \dots, x_N$: the observations
which cannot be expressed analytically in closed form but can be numerically evaluated for any x (Soong, 2004).
If x has the $N(\mu, \sigma)$ distribution, then the standardized variable $z = (x - \mu)/\sigma$ has the standard Normal distribution $N(0, 1)$ with mean 0 and standard deviation 1. Then Z is called the standard variable corresponding to X. In such cases the density function for Z can be obtained from the definition of a Normal distribution by allowing $\mu = 0$ and $\sigma = 1$. The corresponding distribution function is given by (Spiegel *et al*., 2001):

$$F(z) = P(Z \leq z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-u^2/2} \, du \quad (2)$$

where u is a dummy variable of integration.
The method of moments is used to estimate the two parameters of the Normal distribution. According to Chow *et al*. (1988) they are estimated from the following equations:

$$\hat{\mu} = \bar{x} = \frac{\sum_{i=1}^{N} x_i}{N} \quad (3)$$

$$\hat{\sigma} = s = \sqrt{\frac{1}{(N-1)} \sum (x_i - \bar{x})^2} \quad (4)$$

where:
$\bar{x}$: sample mean
s : standard deviation of a sample

2.2 Gamma Distribution

Gamma distribution is widely used in hydrologic analysis. The probability distribution function of a random variable x having Gamma distribution is:

$$F(x) = P(X \le x)$$
$$= \int_0^x f(x)dx = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} \int_0^x x^{\alpha-1} e^{-\frac{x}{\beta}}\, dx \qquad x \ge 0, \alpha > 0, \beta > 0 \tag{5}$$

where:
α : shape parameter
β : scale parameter
$\Gamma(\alpha)$: Gamma function
The two parameter Gamma distribution has a lower bound at zero, which is disadvantage for application to hydrologic variables that have a lower bound larger than zero (Chow, 1988).
Two parameters of Gamma distribution (α and β) are estimated by method of moments as following (Forbes et al., 2011):

$$\alpha = \bar{x}^2/s^2 \tag{6}$$
$$\beta = s^2/\bar{x} \tag{7}$$

2.3 Extreme Value Type III Distribution (Weibull Distribution)

The two parameter Weibull distribution is commonly applied in hydrology since many hydrological variables are bounded by zero (Tallaksen and Lanen, 2004). The cumulative distribution function is given by (Reddy, 1997):

$$F(x) = 1 - e^{(-\frac{x}{\beta})^{\alpha}} \tag{8}$$

where:
α: shape parameter
β: scale parameter
The mean and variance for Weibull distribution are:

$$\mu = \beta\, \Gamma(1 + \frac{1}{\alpha}) \tag{9}$$
$$\sigma^2 = \beta^2 \left[\Gamma\left(1 + \frac{2}{\alpha}\right) - \left[\Gamma\left(1 + \frac{1}{\alpha}\right)\right]^2\right] \tag{10}$$

Since the basic Weibull model has two parameters, estimates of the model parameters can be obtained using the sample mean $\bar{x}$ and sample variance s^2. By equating μ whith $\bar{x}$ and σ^2 with s^2, β is obtained as follows (Murthy et al., 2004):

$$\frac{\sigma^2}{\mu^2} = \frac{s^2}{\bar{x}^2} = \frac{\Gamma(1+\frac{2}{\alpha})}{\Gamma^2(1+\frac{1}{\alpha})} - 1 \tag{11}$$

by comparing Equation (11) with a coefficient of variation equation which is:

$$C_v = \frac{s}{\bar{x}} \tag{12}$$

finding that(Al-Fawzan, 2000):

$$C_v = \frac{\sqrt{\Gamma\left(1+\frac{2}{\alpha}\right) - \Gamma^2\left(1+\frac{1}{\alpha}\right)}}{\Gamma\left(1+\frac{1}{\alpha}\right)} \tag{13}$$

α is obtained by trial and error from the above equation and β is given by the following equation:

$$\beta = \frac{\bar{x}}{\Gamma(1+\frac{1}{\alpha})} \tag{14}$$

3. Goodness of Fit Tests

The goodness of fit tests provides objective procedures to determine whether an assumed theoretical distribution provides an adequate description of the observed data, these tests are valid only for rejecting an inadequate model; they cannot prove that the model is correct (Almog, 1979). Two types are used in this study, these are: Chi-square goodness of fit test, Kolmogorov-Smirnov goodness of fit test.

3.1 Chi-Square Goodness of Fit Test (χ^2)

The chi-square (χ^2) test is a technique that check if a specific distribution of a certain observed event's frequency in a sample is suitable for that sample or not.
The first step in the chi-square test is to arrange the number of observations *N* into a set of *k* cells (class intervals), and then calculate the chi-square statistic from the statistical formula (Guizani, 2010):

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \quad (15)$$

where:
O_i: observed frequency in the *ith* cell
E_i: expected frequency in the same cell
k: number of intervals
The expected frequency can be computed by:

$$E_i = Np_i \quad (16)$$

where

N: total number of observations

p_i: probability of the *ith* cell

the degrees of freedom can be computed using $v = k - s - 1$, where s denotes the number of parameters used in fitting distribution. The null hypotheses of the chi-square test is that there is no difference between observed and estimated value and a tested distribution is describing the observed data, while the alternative hypothesis is that the tested distribution doesn't describe the observed data. The null hypothesis will be rejected if χ^2 is greater than the critical value.

If the distribution being tested is continuous, then a technique called the *equiprobable* approach can be used to divide the data into several cells or class intervals. Each of these cells has the same probability, so that $p_1 = p_2 = \cdots = p_i$, and can be defined as:

$$p_i = \int_{x_{i-1}}^{x_i} f(x)\,dx = F(x_i) - F(x_{i-1}) \quad (17)$$

where x_i and x_{i-1} are the endpoints of the *ith* cell.

If an expected frequency is less than 5, it can be combined with the expected frequency of an adjacent cell, and the corresponding observed frequency should also be combined. Then the value of k should be modified depending on how many cells are combined (Guizani, 2010).

3.2 Kolmogorov-Smirnov goodness of fit test

The Kolmogorov–Smirnov (K-S) test is a goodness-of-fit test used to determine whether an underlying probability distribution differs from a hypothesized distribution when given a finite data set (Guizani, 2010).

The step-by-step procedure for executing K-S test for given a set of sample values $x_1, x_2, \ldots, x_i$ observed from a population X, is as follows (Soong, 2004):

- The sample values are arranged in increasing order of magnitude, denoted by $x_{(i)}$.
- The observed distribution functions $S(x_i)$ are determined from the relation $S(x_i) = i/N$.
- Distribution function $F(x_i)$ at each $x_{(i)}$ by using the hypothesized distribution is obtained and the deviations D_2 are determined from Equation (18) :

$$D_2 = S(x_i) - F(x_i) \quad (18)$$

- The maximum absolute value of D_2, obtained from Equation (18), is compared with critical value shown in statistical tables. If D_2 is less than the critical value the tested distribution is suitable for describing the observed data, otherwise the tested distribution is not suitable for describing the observed data.

4. Statistical Distributions for Iraqi Stations

4.1 Fitting Probability Distributions for Iraqi Stations

Three probability distributions are fitted for thirteen Iraq stations of monthly rainfall (Iraqi Meteorological Office), they are, Normal distribution, Gamma distribution and Weibull distribution. Method of moments is used for parameter estimation as detailed in section two. After parameters estimation, the data must be ordered in increasing way and the cumulative distribution function $F(x)$ is calculated for each. The calculations of distributions for all stations are made using (Minitab 16) software. Table (1) shows the estimated parameter distributions for all Iraqi stations.

Table 1: Estimated parameters for Iraqi stations used in the study

Station	Normal		Gamma		Weibull		
	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\alpha}$	$\hat{\beta}$	C_v	$\hat{\alpha}$	$\hat{\beta}$
Mosul	29.62237	38.81178	0.58252	50.85192	1.31022	0.77200	25.46061
Kirkuk	29.49029	39.80757	0.54882	53.73438	1.34985	0.75150	24.80917
Khanaqin	24.29705	33.85479	0.51507	47.17226	1.39337	0.73050	19.95609
Ramadi	9.08152	13.95473	0.42352	21.44293	1.53661	0.67080	3.64609
Baghdad	10.02090	16.25997	0.37982	26.38351	1.62261	0.64050	7.21375
Hilla	8.27933	13.77694	0.36115	22.92504	1.66402	0.62745	5.81844
Karbala	8.28481	13.99485	0.35045	23.64036	1.68922	0.61980	5.73672
Najaf	8.22765	14.57637	0.31861	25.82396	1.77163	0.59630	5.42426
Diwaniyah	8.95611	15.26094	0.34441	26.00419	1.70397	0.61550	6.14866
Samawa	6.40236	11.43715	0.31336	20.43127	1.78640	0.59250	4.18523
Nasiriyah	10.20387	17.38768	0.34439	29.62908	1.70403	0.61535	7.00318
Amara	13.43099	22.32548	0.36192	37.11024	1.66224	0.62800	9.44872
Basrah	11.56262	18.53916	0.38898	29.72515	1.60337	0.64710	8.42077

4.2 Goodness of Fit Tests for Iraqi Stations

The tests that are used in this study are Chi-Square and Kolmogorov-Smirnov tests.

4.2.1 Chi-Square Test

The chi-square test steps are calculated as mentioned in section 3.1. Table (2) shows the observed and critical Chi-square values for all stations with confidence interval 95%.

Table 2: Observed and critical Chi-Square values for Iraqi stations used in the study

Station	Normal			Gamma			Weibull		
	Observed Chi-square	v	Critical Chi-square	Observed Chi-square	v	Critical Chi-square	Observed Chi-square	v	Critical Chi-square
Mosul	110.9273	9	16.919	49.41983	9	16.919	66.71839	9	16.919
Kirkuk	121.7221	7	14.067	41.4071	7	14.067	55.3233	7	14.067
Khanaqin	132.1982	9	16.919	43.30054	9	16.919	65.25172	8	15.507
Ramadi	84.49457	5	11.07	4.83235	5	11.07	137.13010	2	5.991
Baghdad	157.6434	6	12.592	6.872418	6	12.592	15.92965	5	11.07
Hilla	131.4537	7	14.067	7.5600	7	14.067	18.80558	7	14.067
Karbala	140.48219	8	15.507	16.20691	8	15.507	31.04266	7	14.067
Najaf	183.0854	5	11.07	6.395089	5	11.07	16.76463	5	11.07
Diwaniyah	190.86891	7	14.067	16.3910	7	14.067	28.0693	6	12.592
Samawa	146.08612	6	12.592	12.20760	6	12.592	20.27412	6	12.592
Nasiriyah	169.1617	8	15.507	35.5233	8	15.507	55.9207	7	14.067
Amara	188.1072	8	15.507	20.98796	8	15.507	40.54369	7	14.067
Basrah	197.2412	8	15.507	40.11589	8	15.507	63.14111	8	15.507

From Table (2) it can be seen that Gamma distribution is suitable for five stations: Ramadi, Baghdad, Hilla, Najaf and Samawa stations, Normal and Weibull distributions are not suitable for any station.

4.2.2 Kolmogorov-Smirnov Test

This test, as mentioned in section 3.2 D_2 is calculated and comparing the maximum absolute value of D_2 with the critical value of Kolmogorov-Smirnov test with confidence interval 95% which equals$1.36/\sqrt{N}$. Table (3) shows the final results of Kolmogorov-Smirnov test for Iraq stations used in the study.

Table 3: Observed and critical Kolmogorov-Smirnov values for Iraqi stations used in the study

Station	N	Critical D_2	Observed D_2		
			Normal	Gamma	Weibull
Mosul	492	0.06131	1.0000	0.3049	0.3065
Kirkuk	492	0.06131	0.2274	0.3366	0.3389
Khanaqin	492	0.06131	0.2344	0.3858	0.3895
Ramadi	360	0.07168	0.2548	0.3363	0.3487
Baghdad	492	0.06131	0.2668	0.3647	0.3848
Hilla	372	0.07051	0.2713	0.3518	0.3774
Karbala	492	0.06131	0.2749	0.3802	0.4111
Najaf	492	0.06131	0.2842	0.4073	0.4453
Diwaniyah	492	0.06131	0.2766	0.4215	0.4506
Samawa	492	0.06131	0.2858	0.3709	0.4136
Nasiriyah	492	0.06131	0.2766	0.3864	0.4144
Amara	492	0.06131	0.2717	0.3856	0.4074
Basrah	492	0.06131	0.2644	0.4124	0.4300

From Table (3) it can be seen that not any of the distributions (Normal, Gamma and Weibull) is suitable for the thirteen stations that used in the study.

5. Conclusions

No appropriate distribution is found for the thirteen studied stations because Chi-Square test shows that Gamma distribution is suitable for just five stations: Ramadi, Baghdad, Hilla, Najaf and Samawa stations, Normal and Weibull distributions are not suitable for any station, while Kolmogorov-Smirnov test shows that there is no suitable distribution for any station. So there is no suitable distribution for these stations because one of goodness of fit tests rejected all the tested distributions.

6. References

Al-Fawzan M. A. (2000), “Methods for Estimating the Parameters of the Weibull Distribution”, King Abdulaziz City for Science and Technology.

Al-Mansory H. B. (2005), “Statistical Analysis of Extreme Monthly Rainfall in Basrah City, South of Iraq”, Marina Mesopotamica, Basrah, Iraq, Vol. 20, 2, PP. (283-296).

Almong, R. (1979), “A Study of the Application of the Lognormal Distribution to Corrective Maintenance Repair Time”, M. Sc. Thesis, Naval postgraduate school.

Chow V. T., Maidment D. R. and Mays L. W. (1988), "Handbook of applied hydrology", McGraw-Hill series in water Resources and Environmental Engineering, New York, ISBN 0-07-010810-2.

Forbes C., Evans M., Hastings N. and Peacock B. (2011), “Statistical Distributions”, John Wiley & Sons, Inc., 4th Edition, New Jersey.

Guizani M., Rayes A., Khan B. and Al-Fuqaha A. (2010), “Network Modeling and Simulation: a Practical Perspective”, John Wiley & Sons Ltd., U. S. A.

Husak G. J., Michaelsen J. and Funk C. (2006), “Use of the Gamma Distribution to Represent Monthly Rainfall in Africa for Drought Monitoring Applications”, International Journal of Climatology, Published online in Wiley Inter. Science, Royal Meteorological Society, Santa Barbara, U. S. A., DOI: 10.1002/joc.

Iraqi Meteorological Office, Baghdad, Iraq.

Maliva R. and Missimer T. (2012), “Arid Lands and Water Evaluation and Management”, Springer-Verlag Berlin Heidelberg, ISBN 978-3-642-29104-3.

Murthy D. P., Xie M. and Jiang R. (2004), “Weibull Models”, John Wiley & Sons, Inc., New Jersey.

Olumide B. A., Saidu M., Oluwasesan A. (2013), “Evaluation of Best Fit Probability Distribution Models for the Prediction of Rainfall and Runoff Volume (Case Study Tagwai Dam, Minna-Nigeria)”, International Journal of Engineering and Technology, U. K., Vol. 3, No. 2, PP. (94-98), ISSN: 2049-3444.

Ozturk A. (1981), “On the Study of a Probability Distribution for Precipitation Totals”, Department of Statistics, Division of Mathematical Science, Mimeograph Series (81-3).

Papalexiou S. M., Koutsoyiannis D., and Makropoulos C. (2012), “How Extreme is Extreme? An Assessment of Daily rainfall Distribution Tails”, Discussion Paper, Journal of Hydrology and Earth System Sciences, Vol. 9, PP. (5757–5778).

Reddy J. R. (1997), “Stochastic Hydrology”, Laxmi Publications, 2nd Edition, New Delhi, India, ISBN 10: 817008086X.

Sen Z. and Eljadid A. G. (1999), “Rainfall Distribution Function for Libya and Rainfall Prediction”, Hydrological Sciences Journal, Istanbul, Turkey, Vol.44, 5, PP. (665-680), Posted on internet website http://dx.doi.org/10.1080/02626669909492266

Soong T.T. (2004), "Fundamentals of Probability and Statistics for Engineers", John Wiley & Sons Ltd, State University of New York at Buffalo, Buffalo, New York, USA.

Spiegel, M. R., Schiller, J., and Srinivasan R. A. (2001), "Based on Schaum’s Outline of probability and statistics", Schaum’s Outline series, Mcgraw-Hill, New Yourk.

Tallaksen L. M. and Lanen H. V. (2004), “Hydrological Drought: Processes and Estimation Methods for Streamflow and Groundwater”, Elsevier.